



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2010

A collaborative system for sentiment analysis

Rentoumi, V ; Petrakis, S ; Karkaletsis, V ; Klenner, M ; Vouros, G A

DOI: https://doi.org/10.1007/978-3-642-12842-4_52

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-39616>

Conference or Workshop Item

Originally published at:

Rentoumi, V; Petrakis, S; Karkaletsis, V; Klenner, M; Vouros, G A (2010). A collaborative system for sentiment analysis. In: 6th Hellenic Conference on Artificial Intelligence (SETN 2010), Athens, Greece, 4 May 2010 - 7 May 2010.

DOI: https://doi.org/10.1007/978-3-642-12842-4_52

A Collaborative System for Sentiment Analysis

Vassiliki Rentoumi^{1,2}, Stefanos Petrakis³, Vangelis Karkaletsis¹,
Manfred Klenner³, George A. Vouros²

¹Inst. of Informatics and Telecommunications, NCSR “Demokritos”, Greece

²University of the Aegean, Artificial Intelligence Laboratory, Samos, Greece

³Institute of Computational Linguistics, University of Zurich, Switzerland

vrentoumi@iit.demokritos.gr petrakis@cl.uzh.ch vangelis@iit.demokritos.gr
klenner@cl.uzh.ch georgev@aegean.gr

Abstract. In the past we have witnessed our machine learning method for sentiment analysis coping well with figurative language, but determining with uncertainty the polarity of mildly figurative cases. We have shown that for these uncertain cases, a rule-based system should be consulted. We evaluate this collaborative approach on the “Rotten Tomatoes” movie reviews dataset and compare it with other state-of-the-art methods, providing further evidence in favor of this approach.

1 Introduction

In the past we have shown that figurative language conveys sentiment that can be efficiently detected by FigML[2], a machine learning (ML) approach trained on corpora manually annotated with strong figurative expressions¹. FigML was able to detect the polarity of sentences bearing highly figurative expressions, where disambiguation is considered mandatory, such as: (a) “credibility sinks into a mire of sentiments”. On the other hand, there exist cases for which FigML provided a classification decision based on a narrow margin between negative and positive polarity orientation, often resulting in erroneous polarity evaluation. It was observed that such cases bear mild figurativeness, which according to [4] are synchronically as literal as their primary sense, as a result of standardized usage, like: (b) “this 10th film in the series looks and feels tired”. Here, fatigue as a property of inanimate or abstract objects, although highly figurative, presents an obvious negative connotation, due to standardized usage of this particular sense, therefore sentiment disambiguation is not necessary. Such regular cases could be more efficiently treated by a rule-based system such as PolArt[1].

In fact, in this paper we extend the work presented in [8] where we have indeed shown that cases of mild figurative language are better treated by PolArt, while cases of strong figurative language are better handled by FigML. In [8], a novel collaborative system for sentiment analysis was proposed and managed

¹ Subsets from the AffectiveText corpus (SemEval’07) and the MovieReviews sentence polarity dataset v1.0, annotated with metaphors and expanded senses: <http://www.iit.demokritos.gr/~vrentoumi/corpus.zip>

to outperform its two subcomponents, FigML and PolArt, tested on the AffectiveText corpus. Here, we try to verify the validity of this approach on a larger corpus and of a different domain and style. In addition and most importantly, another dimension of complementarity between a machine learning method and a rule-based one is explored: the rule-based approach handles the literal cases and the - already introduced - collaborative method treats the cases of figurative language. Results show that integrating a machine learning approach with a finer-grained linguistically-based one leads to a superior, best-of-breed system.

2 Methodology Description

The proposed collaborative method involves four consecutive steps:

(a) Word sense disambiguation (WSD): We chose an algorithm which takes as input a sentence and a relatedness measure [6]. The algorithm supports several WordNet based similarity measures among which Gloss Vector (GV) [6] performs best for non-literal verbs and nouns [5]. Integrating GV in the WSD step is detailed in [2].

(b) Sense level polarity assignment (SLPA): We adopted a machine learning approach which exploits graphs based on character n-grams [7]. We compute models of positive and negative polarity from examples of positive and negative words and definitions provided by an enriched version of the Subjectivity Lexicon². The polarity class of each test sense, is determined by computing its similarity with the models as detailed in [2].

(c) HMMs training: HMMs serve two purposes. Computing the threshold which divides the sentences in marginal/non-marginal and judging the polarity (positive/negative) of non-marginal sentences. We train one HMM model for each polarity class. The format of the training instances is detailed in [2]. For computing the threshold, the training data are also used as a testing set.

Each test instance is tested against both models and the output is a pair of log probabilities of a test instance to belong to either the negative or the positive class. For each polarity class we compute the absolute difference of the log probabilities. We then sort these differences in ascending order and calculate the first Quartile (Q1) which separates the lower 25% of the sample population from the rest of the data. We set this to be the threshold and we apply it to the test instances. Marginal cases are the ones for which the absolute difference of log probability is below that threshold. In our experiments we use a 10-fold cross validation approach to evaluate our results.

(d) Sentence-level polarity detection: The polarity of each sentence is determined by HMMs [2] for non-marginal cases and by PolArt [1] for marginal ones. PolArt employs compositional rules and obtains word-level polarities from

² <http://www.cs.pitt.edu/mpqa/>

³ For each positive or negative word entry contained in the Subjectivity Lexicon, we extracted the corresponding set of senses from WordNet, represented by their synsets and gloss examples; in this way we tried to reach a greater degree of consistency between the test and the training set

a polarity lexicon, as described in detail in [1]. The Collaborative system’s total performance is then given by adding up the performances of FigML and PolArt.

3 Experimental Setup

3.1 Resources

We ran our experiments on the MovieReviews corpus⁴. This corpus was split into different subsets according to our experimental setup in two different ways:

- Expanded Senses/Metaphors/Whole: The corpus was enriched with manually-added annotations for metaphorical and expanded senses expressions inside sentences. We produced an expanded senses dataset and a metaphorical expressions one. Furthermore, we treated the entire corpus as a third dataset, ignoring the aforementioned annotations. The produced datasets are:
 - Expanded senses: 867 sentences, 450 negative and 417 positive ones.
 - Metaphors: 996 sentences, 505 negative and 491 positive ones.
 - Whole: 10649 sentences, 5326 negative and 5323 positive ones.
- Literal/Non-literal: We group all figurative sentences (metaphors/expanded senses) as the non-literal set. The rest of the sentences we call the literal set.
 - Non-literal: 1862 sentences⁵, 954 negative and 908 positive ones.
 - Literal: 8787 sentences, 4372 negative and 4415 positive ones.

We run numerous variations of PolArt, modifying each time the polarity lexicon it consults:

- SL+: This is the subjectivity lexicon⁶ with manually added valence operators.
- Merged: The FigML system produces automatically sense-level polarity lexica (AutSPs), one for each dataset or subset. For the non-literal, metaphors and expanded senses, these lexica target non-literal expressions, metaphors and expanded senses accordingly. For the entire MovieReviews dataset (Whole), all word senses are targeted. Various Merged lexica are produced by combining and merging the SL+ lexicon with each of the AutSPs.

3.2 Collaborative Method Tested on MovieReviews dataset

We tested our Collaborative method originally presented and evaluated in [8], with the extended MovieReviews corpus, in order to test its validity. Table 1 presents scores for each polarity class, for both variants of the our method, the CollaborativeSL+ (using the SL lexicon) and CollaborativeMerged (using the

⁴ We used the sentence polarity dataset v1.0 from <http://www.cs.cornell.edu/People/pabo/movie-review-data/>

⁵ One sentence belonged to both the metaphors and expanded senses subsets, and was included only once here

⁶ <http://www.cs.pitt.edu/mpqa/>

Merged Lexica), across all three datasets. For the majority of cases, CollaborativeSL+ has better performance than CollaborativeMerged. Comparing the performance of CollaborativeSL+ for the MovieReviews with that of CollaborativeSL+ for the AffectiveText corpus [8], for the Whole corpus (f-score:neg pos), we noticed that the performance remains approximately the same. This is evidence that the method is consistent across different datasets.

		CollaborativeSL+		CollaborativeMerged	
		neg	pos	neg	pos
Whole	recall	0.682	0.537	0.656	0.536
	precision	0.596	0.628	0.586	0.609
	f-measure	0.636	0.579	0.619	0.570
Met	recall	0.724	0.735	0.697	0.704
	precision	0.737	0.722	0.708	0.693
	f-measure	0.731	0.728	0.702	0.699
Exp	recall	0.640	0.623	0.642	0.623
	precision	0.647	0.616	0.648	0.617
	f-measure	0.643	0.619	0.645	0.620

Table 1. MovieReviews: Performance scores for full system runs

3.3 The collaborative approach treats non-literal cases as a whole : complementarity on the literal/non-literal axis

We have so far shown that our Collaborative method is performing quite well on the expanded senses and metaphors datasets. Although we consider them as distinct language phenomena, they both belong to the sphere of figurative connotation. To support this we tested our claim collectively, across non-literal expressions in general, by merging these two datasets into one labelled non-literals. As a baseline system for assessing the performance of the collaborative method we use a clean version of PolArt (i.e. without added valence shifters).

In Table 2, we compare BaselinePolart with CollaborativeSL+ (using the SL lexicon) and CollaborativeMerged (using the Merged Lexica), tested upon the non-literals dataset. We observe that our proposed method outperforms the baseline and proves quite capable of treating non-literal cases collectively.

		CollaborativeSL+		CollaborativeMerged		BaselinePolart	
		neg	pos	neg	pos	neg	pos
Nonliterals	recall	0.710	0.646	0.681	0.644	0.614	0.667
	precision	0.678	0.680	0.668	0.658	0.659	0.622
	f-measure	0.694	0.662	0.674	0.651	0.636	0.644

Table 2. MovieReviews: Performance scores for the non-literals subset

By assembling the non-literals into one dataset and treating it with our collaborative method we set aside its complementary dataset of literals. Since our

method is more inclined to treat figurative language, we do not expect that it should treat literal cases optimally, or at least as efficiently as a system that is more inclined to treat literal language. Therefore, assigning the literals to PolArt and the nonliterals to Collaborative, would provide a more sane system architecture and result in better performance for the entire MovieReviews dataset.

		PolartwithCollaborativeSL+		PolartwithCollaborativeMerged	
		neg	pos	neg	pos
Literals/nonliterals	recall	0.608	0.659	0.603	0.659
	precision	0.641	0.627	0.638	0.624
	f-measure	0.624	0.642	0.620	0.641
		CollaborativeSL+		CollaborativeMerged	
		neg	pos	neg	pos
Whole	recall	0.682	0.537	0.656	0.536
	precision	0.596	0.628	0.586	0.609
	f-measure	0.636	0.579	0.619	0.570

Table 3. MovieReviews: Performance scores for full system runs

In Table 3 we present the performance of both variants of the new system architecture (PolartwithCollaborativeSL+, PolartwithCollaborativeMerged). In both versions pure PolArt treats literal cases, while CollaborativeSL+ and CollaborativeMerged treat non literals cases. This new architecture is compared to the one concerning the treatment of the whole corpus (Whole) by both variants of the proposed method (CollaborativeSL+, CollaborativeMerged). It is observed that the performance of this modified system is better for the majority of cases. This fact leads us to the conclusion that a system which treats sentiments in a more language-sensitive way, can exhibit improved performance.

We further compared our system with a state-of-the-art system by Andreevskaia and Bergler[3], tested on the MovieReviews corpus. Their system employs a Naive Bayes Classifier for polarity classification of sentences, trained with unigrams, bigrams or trigrams derived from the same corpus. This state-of-the-art system’s accuracy was reported to be 0.774, 0.739 and 0.654 for unigrams, bigrams and trigrams. Our two alternative system architectures, CollaborativeSL+ and PolartwithCollaborativeSL+, scored 0.609 and 0.633.

The performances of both our alternatives are clearly lower than the state-of-the-art system’s when the latter is trained with unigrams or bigrams, but they get closer when it is trained with trigrams. The main point is that the CollaborativeSL+ method performs quite well even for the case of a corpus containing mainly literal language. We expect CollaborativeSL+ to perform optimally when applied on a corpus consisting mainly of non-literal language.

It is also worth noting that since PolArt deals with the majority of cases it is bound to heavily affect the overall system performance. Additionally PolArt’s dependency on its underlying resources and especially the prior polarity lexicon is also a crucial performance factor. Thus, the observed moderate performance of the system can be attributed to the moderate PolArt’s performance,

probably due to the incompatibility of the Subjectivity Lexicon with the idiosyncratic/colloquial language of the Movie Reviews corpus.

All in all, the overall performance is still quite satisfactory. Consequently, if we provide PolArt with a more appropriate lexicon, we expect a further boost.

4 Conclusions and Future Work

In this paper we further extend and examine the idea of a sentiment analysis method which exploits complementarily two language specific subsystems, a rule-based (PolArt) for the mild figurative, and a machine learning system (FigML) for the strong figurative language phenomena[8]. By further examining the validity of such an approach in a larger (and of different domain) corpus (Movie Reviews corpus), in which strong figurative language co-exists with mild figurative language, we observed that this Collaborative method is consistent.

We also explored another dimension of complementarity concerning literal/non-literal cases of language, where PolArt is treating the literal cases and the Collaborative method the non-literal cases. We get empirical support from the performance obtained that utilizing the special virtues of the participating subsystems can be a corner-stone in the design and performance of the resulting system.

We will test the collaborative method on a more extensive corpus bearing figurative language. We intend to dynamically produce sense-level polarity lexica exploiting additional machine learning approaches (e.g. SVMs).

References

1. M. Klenner, S. Petrakis, and A. Fahrni. Robust compositional polarity classification. Recent Advances in Natural Language Processing (RANLP), Borovets, Bulgaria, 2009.
2. V. Rentoumi, G. Giannakopoulos, V. Karkaletsis, and G. Vouros. Sentiment analysis of figurative language using a word sense disambiguation approach. Recent Advances in Natural Language Processing (RANLP), Borovets, Bulgaria, 2009.
3. A. Andreevskaia and S. Bergler. When specialists and generalists work together: overcoming domain dependence in sentiment tagging. Proceedings of ACL-08: HLT, pages:290–298, 2008.
4. D.A. Cruse. Meaning in language Oxford University Press, 2000
5. V. Rentoumi, V. Karkaletsis, G. Vouros and A. Mozer. Sentiment Analysis Exploring Metaphorical and Idiomatic Senses: A Word Sense Disambiguation Approach. International Workshop on Computational Aspects of Affectual and Emotional Interaction (CAFEEi 2008).
6. T. Pedersen, S. Banerjee and S. Patwardhan. Maximizing Semantic Relatedness to Perform Word Sense Disambiguation. Supercomputing Institute Research Report UMSI, v. 25, 2005.
7. G. Giannakopoulos, V. Karkaletsis, G. Vouros and P. Stamatopoulos. Summarization system evaluation revisited: N-gram graphs. ACM Transactions on Speech and Language Processing (TSLP) ACM New York, NY, USA, v: 5, 2008
8. V. Rentoumi, S. Petrakis, M. Klenner, G. Vouros and V. Karkaletsis. A Hybrid System for Sentiment Analysis. To appear in LREC 2010.